

2006
STATISTICS

Paper 1

Time : 3 Hours]

[Maximum Marks : 300

INSTRUCTIONS

*Candidates should attempt **all** the questions in Parts A, B & C. However, they have to choose only **three** questions in Part D. The number of marks carried by each question is indicated at the end of the question.*

Answers must be written in English.

This paper has four parts :

A	20 marks
B	100 marks
C	90 marks
D	90 marks

Marks allotted to each question are indicated in each part.

Assume suitable data if considered necessary and indicate the same clearly.

Notations and symbols used are as usual.

SEAL

PART A

4×5=20

Each question carries 5 marks.

1. (a) If $\{A_n\}$, $n \geq 1$ is a sequence of events for which limit exists, then prove that

$$\lim_n P(A_n) = P(\lim_n A_n)$$

- (b) If X_1 and X_2 are independent and identically distributed $B(1, p)$ random variables, then examine whether $(X_1 + 2X_2)$ is sufficient for p .
- (c) Distinguish between simple and composite hypotheses. Illustrate.
- (d) Explain the practical advantages of a sequential test procedure.

PART B

10×10=100

Each question carries 10 marks.

2. (a) Stating the regularity conditions on $f(x, \theta)$, $\theta \in \Omega$, write down the Cramer – Rao lower bound for the variance of an unbiased estimator of $g(\theta)$.
- (b) Let $X \sim N(0, 1)$. Based on a random sample X_1, \dots, X_n , construct an unbiased estimator of θ^2 . Also obtain the Cramer – Rao lower bound on variance. Is this bound attained in this case ?
3. (a) Let $f(x, y) = \frac{1}{4} [1 - x^3y]$, $-1 < x, y < 1$ be the joint density function. Derive the marginal density functions of X and Y .
- (b) Specify the probability density function (pdf) of a bivariate normal distribution. In the case of zero correlation show that the components are independently distributed.
4. Establish Holder's inequality and deduce Cauchy – Schwartz inequality as a particular case.
5. State Lindberg – Feller central limit theorem. Examine whether this result holds for the sequence $\{X_n\}$ of independent random variables with

$$P(X_n = -n) = \frac{1}{2\sqrt{n}} = P(X_n = n) \quad \text{and} \quad P(X_n = 0) = 1 - \frac{1}{\sqrt{n}}.$$
6. Given the characteristic function $\phi(t) = \exp\{-|t|\}$, show that the pdf exists and obtain the same.
7. Establish Lehmann – Scheffé theorem. State the significance of this result.
8. Outline the chi-square test for goodness of fit. Mention clearly the underlying assumptions.
9. Define the p-variate normal distribution. Derive its characteristic function.

[Turn over

10. Illustrate the application of the following :

- (a) Hotellings' T^2 statistic
- (b) Mahalanobis' D^2 statistic

11. Stating the necessary conditions on X and U , show that the least squares estimator of β in the model $Y = X\beta + U$ is (a) unbiased and (b) consistent.

PART C

6×15=90

Each question carries 15 marks.

12. (a) Define sufficient, minimal sufficient and complete minimal sufficient statistics.
- (b) Define Fisher's information function and compute it for the one-parameter exponential family of density functions.
13. (a) Define a maximum likelihood estimator (MLE) and state its important properties.
- (b) Obtain the MLE of θ based on random samples of n observations from a uniform distribution over the interval $(0, \theta + 1)$.
14. Outline the following methods of estimation :
- (a) Method of moments
- (b) Method of minimum chi-square
15. Establish Markov inequality and hence deduce Chebychev's inequality.
16. (a) Derive the distribution of the sample multiple correlation when the population multiple correlation is zero in the trivariate case.
- (b) Define orthogonal polynomials. How are these useful in regression analysis ?
17. Explain the following for a two-sample problem :
- (a) Wilcoxon – Mann – Whitney test
- (b) Kolmogorov – Smirnov test

[Turn over

PART D

3×30=90

Answer any **three** of the following questions. Each question carries 30 marks.

18. (a) State and prove Khinchine's weak law of large numbers (WLLN).

(b) Examine whether the WLLN holds for the sequence of random variables $\{X_n\}$ with $P(X_n = n) = \frac{1}{2n\sqrt{n}} = P(X_n = -n)$ and

$$P(X_n = 0) = 1 - \frac{1}{n\sqrt{n}}.$$

(c) If $\{X_n\}$ is a sequence of independent random variables with

$$P(X_n = 0) = 1 - \frac{1}{n} = 1 - P(X_n = 1),$$

using Borel – Cantelli lemma, prove that $X_n \rightarrow 0$ almost surely, as $n \rightarrow \infty$.

19. State and prove Neyman – Pearson fundamental lemma. Explain clearly the role of this result in tests of hypotheses.

20. Establish the Gauss – Markov theorem. What are the implications of this theorem for regression analysis? Elaborate.

21. (a) State the advantages with non-parametric tests.

(b) Explain the assumptions and operating procedure of the following :

(i) Runs test

(ii) Sign test

(iii) Construction of confidence intervals for quantiles

22. (a) Explain the technique of Analysis of Variance. When and how is this technique useful? Illustrate.

(b) Giving an example, state the objective of discriminant analysis. Define Fisher's linear discriminant function and explain its use.

SEAL

2006
STATISTICS

Paper 2

Time : 3 Hours]

[Maximum Marks : 300

INSTRUCTIONS

Candidates should attempt **all** the questions in Parts A, B & C. However, they have to choose only **three** questions in Part D. The number of marks carried by each question is indicated at the end of the question.

Answers must be written in English.

This paper has four parts :

A	20 marks
B	100 marks
C	90 marks
D	90 marks

Marks allotted to each question are indicated in each part.

Assume suitable data if considered necessary and indicate the same clearly.

Notations and symbols used are as usual.

SEAL

PART A

4×5=20

Each question carries 5 marks.

1. (a) Define linear systematic sampling from a finite population. State the advantages and limitations of the scheme.
- (b) Explain the significance of the randomization principle while designing field experiments.
- (c) Describe a linear programming problem (LPP). Illustrate.
- (d) Explaining the notations, give the formula for Fisher's price index. Why is this index referred to as "ideal" ?

PART B

10×10=100

Each question carries 10 marks.

2. Let y denote the study variate and x , a related auxiliary variate. Assume that information on x is available. Explain how this information can be used for
 - (a) forming ratio estimators,
 - (b) as size measure in PPS sampling.
3. Explain clearly the missing plot technique with reference to a Randomized Block Design (RBD). How does this affect the error degrees of freedom ?
4. Describe the operation of a double sampling plan for acceptance sampling by attributes. Derive the OC and ASN functions.
5. Evaluate the reliability of the k -component standby system, where each component obeys exponential distribution with parameter θ .
6. Describe an inventory situation where the demand is stochastic in nature. Outline the method of deriving an optimal policy for this case.
7. Give two examples for time series data. Describe, with necessary notations, the least squares method for linear trend fitting.
8. How is a consumer price index for industrial workers constructed ? Explain. Can this be used to describe the purchasing power of money ?
9. Mention the common sources of multicollinearity in regression analysis. How is this problem handled ? Explain.
10. Define the following measures and state their use :
 - (a) Infant Mortality Rate
 - (b) Sex Ratio
 - (c) Net Reproduction Rate
11. Explaining the different columns, describe the method of constructing a life-table.

[Turn over

PART C

6×15=90

Each question carries 15 marks.

12. When and how is stratification of a population before sampling useful ?
In usual notation in stratified random sampling show that

$$V_{\text{opt}} \leq V_{\text{prop}} \leq V_{\text{ran}}$$

What are the implications of this result ?

13. (a) Explain the characteristics of a queuing system. Illustrate with an example.
- (b) For the $M|M|\infty$ system show that the number of customers in the system at time t is distributed as Poisson.
14. Describe a CUSUM control chart. Compare this chart with a Shewhart chart, with reference to performance. What is a V-mask and how are the components of this mask determined ?
15. (a) Establish the relation between the hazard and reliability functions of a component.
- (b) Examine whether Weibull and Pareto distributions are IFR distributions.
16. Explain transportation and assignment problems. Are these particular cases of a LPP ? Justify your answer.
17. (a) Describe the logistic growth model. State its chief characteristics.
- (b) How is logistic model fitted to population data ? Explain.

PART D

3×30=90

Answer any **three** of the following questions. Each question carries 30 marks.

18. (a) Define a BIBD. In the usual notations show that for a BIBD,
 $b \geq v + r - k$.
- (b) Briefly explain the following :
- (i) Confounding of an effect
 - (ii) Split plot experiments
19. (a) Explain the construction, basis and use of a p-chart.
- (b) What are rectification plans for acceptance sampling by attributes ? Explain the meaning and significance of the AOQ and AOQL concepts.
20. Describe the simplex algorithm for solving a LPP. What is Charnes M method ? Explain.
21. Specify the identification problem in simultaneous equation models. Derive the rank and order conditions for identifiability. Interpret the result.
22. Write briefly on the following :
- (a) Input devices for a computer
 - (b) Supply and demand elasticities
 - (c) T and Z scores in educational tests